ABSTRACT OF THE DISCLOSURE

A single high resolution frame is obtained from a plurality of frames sampled from moving image data, regardless of movement of a subject included in the frames. Patches are provided in a plurality of frames including a reference frame and other frames. The patches are moved and/or deformed, and correspondent relationships are estimated for each frame. The other frames are coordinate converted based on the estimated correspondent relationships. Correlative values, which represent the degrees of correlation between the other frames and the reference frame, are calculated. The number of regions within the patches is changed, and the correlative values are calculated in the same manner. A synthesized frame is generated from the plurality of frames based on the correspondent relationship, which was estimated for the number of regions that yields the maximal correlation.